DOCUMENT RESUME

ED 442 867 TM 031 287

AUTHOR Pan, Wei; Frank, Kenneth A.
TITLE On the Moments of the Product of Two Dependent Correlation Coefficients.
PUB DATE 2000-04-00
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Correlation; *Sample Size; Simulation; Social Science Research
IDENTIFIERS *Kurtosis; Mean (Statistics); *Product Moment Correlation Coefficient; Skew Curves; Variance (Statistical)

ABSTRACT

A simulation study was conducted to obtain approximate moments of the distribution of the product of two dependent correlation coefficients. The parameters in the simulation were sample size and three population corrections. In general, the simulation results were not very satisfactory. The approximated mean, variance, and skewness were accurate enough, but the kurtosis was not. The implications of these findings for other social science research are discussed. (Contains 1 table and 35 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Wei Pan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 442 867

On the Moments of the Product of Two Dependent Correlation Coefficients

Wei Pan

Michigan State University

Kenneth A. Frank

Michigan State University

April 26, 2000

Direct correspondence to:
Wei PAN
Measurement and Quantitative Methods
College of Education
Michigan State University
East Lansing, MI 48824-1034
E-mail: panwei@pilot.msu.edu
Phone: (517) 355-2893

This paper was prepared for the AERA annual meeting in New Orleans, April 24-28, 2000.

TM031287

BEST COPY AVAILABLE

Introduction

This work is motivated by Frank (in press), in which an index, a product of two dependent correlation coefficients, is derived for assessing the impact of a confounding variable on a regression coefficient. People are very concerned about the impact of a confounding variable, when they make casual inferences from statistical linear models. This is because in most social sciences we are usually not able to conduct a randomized experiment. Thus, if we also cannot statistically control for confounding variables, we will not be able to make a valid casual inference from the results of the statistical analysis (Abbott, 1998; Cook & Campbell, 1979; Sobel, 1996).

If a confounding variable were introduced into the linear model, the effect of the predictor of interest on the outcome could be changed from significant to insignificant. The impact of the confounding variable can be obtained by re-expressing a *t*-value as

$$t = \frac{\hat{\beta}_x}{se(\hat{\beta}_x)} = \frac{r_{xy} - r_{xc} r_{yc}}{\sqrt{\frac{1 - r_{xy}^2 - r_{xc}^2 - r_{yc}^2 + 2 r_{xy} r_{xc} r_{yc}}{n - q - 1}}}. \qquad (1)$$

If the coefficient β_x changes from statistically significant to insignificant, the *t*-value (1) crosses the threshold of the critical value. Here

$se(\hat{\beta}_x)$ is the standard error of $\hat{\beta}_x$;

$\hat{\beta}_x$ is an estimate of the regression coefficient of X;

X is the predictor of interest;

Y is the outcome;

C is the confounding variable;

n is the sample size;

q is the number of independent variables, e.g., X and C; and

r_{xy}, r_{xc}, and r_{yc} are the correlation coefficients between X and Y, between X and C, and between Y and C, respectively.

For example, using a simple regression we found in the NELS:88 data (NCES, 1992) that *parent control* has a significant effect ($p < .025$) on *students' reading achievement*. Also, numerous researches have shown that *students' socioeconomic status* is related to both *parent control* and *students' reading achievement*. Thus, *students' socioeconomic status* is a potential confounding variable for *parent control* on *students' reading achievement*. By introducing *students' socioeconomic status* into the simple regression, we found that the effect of *parent control* on *students' reading achievement* becomes insignificant ($p > .217$), proving that *students' socioeconomic status* is a confounding variable for *parent control* on *students' reading achievement*.

To quantify the impact of a confounding variable, Frank (in press) defines the impact as a product of the two correlations r_{xc} and r_{yc}: $k = r_{xc} \times r_{yc}$. Given this constraint, the t-value (1) achieves its minimum value when $r^2_{xc} = r^2_{yc} = k$. Thus, the t-value can be re-expressed as

$$t_{\min} = \frac{r_{xy} - k}{\sqrt{\frac{(1 + r_{xy} - 2k)(1 - r_{xy})}{n - q - 1}}}, \tag{2}$$

which affords the confounding variable the greatest impact on the inference regarding the regression coefficient of X. This expression reveals the relationship between the index k and t-value. That is, we can express k as a function of estimated correlation r_{xy} and the minimum t-value $t_{\min}$ (2). Therefore, if we set the minimum t-value equal to a critical value, then we can obtain the magnitude of k which is necessary for a confounding variable to alter the inference of the effect of the predictor on the outcome, even if we are not able to measure the confounding variable. In other words, if we had a confounding variable, we could know from the index k how strong the relationships (correlations) of the confounding variable with the predictor and with the outcome should be to alter the interpretation of the predictor.

In practice, Frank (in press) suggests generating a reference distribution for the impact of an unmeasured confounding variable from the impact of existing, measured covariates. The impact of a covariate is defined as the product of two correlations between the predictor of interest and the covariate and between the outcome and the covariate.

From the structural property of the problem in Frank (in press), it is quite likely that there may be a number of such problems in other social science disciplines. Therefore, it is desirable to know the behavior of the distribution of the product of two dependent correlations. From Cohen and Cohen (1983, p. 280), we know that the product of two de-

pendent correlations is constrained by the upper and lower limits, rather than just - 1 and 1:

$$r_{xy} - \sqrt{(1-r_{xc}^2)(1-r_{yc}^2)} < r_{xc}r_{yc} < r_{xy} + \sqrt{(1-r_{xc}^2)(1-r_{yc}^2)}\,. \qquad (3)$$

Thus, the distribution of the product is odd, obviously not normally distributed. The purpose of this study is to obtain approximate moments of the distribution of the product of two dependent correlation coefficients.

Prior Research

We know from the asymptotic distribution theory that the distribution of a set of correlation coefficients approaches multivariate normal, as the sample size becomes large enough (Steiger, 1980, p.246). Hence, the aforementioned problem could be solved by a straightforward approach that treats the product of the two correlations as a product of two normal variables asymptotically and incorporates the findings in the literature about the distribution of the product of two normal variables (Aroian, 1947; Aroian, Taneja, & Cornwell, 1978; Cornwell, Aroian, & Taneja, 1978; Craig, 1936, 1942; Meeker, Cornwell, & Aroian, 1981; Meeker & Escorbar, 1994; Springer, 1983; Wallgern, 1980). However, the convergence of this approach is very slow (Craig, 1936). Thus, this approach presents a severe limitation in application.

There are two other plausible approaches to get the distribution of the product of two correlations. Mathai and Saxena (1969) express the product of two correlations as a special case of the product of two generalized Mellin-Barnes functions or *H*-functions (Mathai & Saxena, 1978). However, the expression obtained for the distribution function is also quite unwieldy. In addition, in their study, the two correlations are assumed independent, while the two relevant correlations defining the impact index *k* are associated with a common variable—the confounding variable or covariate. Therefore, the two correlations that we are interested in are not statistically independent (Boyer, Palachek, & Schucany, 1983; Choi, 1977; Dunn & Clark, 1969, 1971; May & Hittner, 1997a, b; Meng, Rosenthal, & Rubin, 1992; Neill & Dunn, 1975; Steiger, 1980; Williams, 1959; Wolfe, 1976).

The other approach is described in Frank (in press). He transforms the two correlations to two normally distributed Fisher-*z*'s, then uses the findings about the distribution of the product of two normal variables to get the distribution of the product of two Fisher-*z*'s, instead of two original correlations. A problem with this approach, however, is that we are interested in the product of two original correlations and that the behavior of the distribution of the product of two Fisher-*z*'s is very different from that of the distribution of the product of two original correlations.

Approximation Procedures

Let ρ_{xy}, ρ_{xc}, and ρ_{yc} be the population values of r_{xy}, r_{xc}, and r_{yc}, respectively. Set $r_{xy} - \rho_{xy} = \Delta r_{xy}$, $r_{xc} - \rho_{xc} = \Delta r_{xc}$, and $r_{yc} - \rho_{yc} = \Delta r_{yc}$, the deviations of the sample values from their population values. Then, we have

$$r_{xc}r_{yc} = \rho_{xc}\rho_{yc} + \rho_{xc}\Delta r_{yc} + \rho_{yc}\Delta r_{xc} + \Delta r_{xc}\Delta r_{yc};$$

$$\begin{aligned}(r_{xc}r_{yc})^2 = {} & \rho^2_{xc}\rho^2_{yc} + \rho^2_{xc}(\Delta r_{yc})^2 + \rho^2_{yc}(\Delta r_{xc})^2 + (\Delta r_{xc})^2(\Delta r_{yc})^2 + 2\rho^2_{xc}\rho_{yc}\Delta r_{yc} \\ & + 2\rho_{xc}\rho^2_{yc}\Delta r_{xc} + 2\rho_{yc}(\Delta r_{xc})^2\Delta r_{yc} + 2\rho_{xc}\Delta r_{xc}(\Delta r_{yc})^2 + 4\rho_{xc}\rho_{yc}\Delta r_{xc}\Delta r_{yc};\end{aligned}$$

$$\begin{aligned}(r_{xc}r_{yc})^3 = {} & \rho^3_{xc}\rho^3_{yc} + \rho^3_{xc}(\Delta r_{yc})^3 + \rho^3_{yc}(\Delta r_{xc})^3 + (\Delta r_{xc})^3(\Delta r_{yc})^3 + 3\rho^3_{xc}\rho^2_{yc}\Delta r_{yc} \\ & + 3\rho^3_{xc}\rho_{yc}(\Delta r_{yc})^2 + 3\rho^2_{xc}\rho^3_{yc}\Delta r_{xc} + 3\rho_{xc}\rho^3_{yc}(\Delta r_{xc})^2 + 3\rho^2_{xc}\Delta r_{xc}(\Delta r_{yc})^3 \\ & + 3\rho_{xc}(\Delta r_{xc})^2(\Delta r_{yc})^3 + 3\rho^2_{yc}(\Delta r_{xc})^3\Delta r_{yc} + 3\rho_{yc}(\Delta r_{xc})^3(\Delta r_{yc})^2 \\ & + 9\rho^2_{xc}\rho^2_{yc}\Delta r_{xc}\Delta r_{yc} + 9\rho^2_{xc}\rho_{yc}\Delta r_{xc}(\Delta r_{yc})^2 + 9\rho_{xc}\rho^2_{yc}(\Delta r_{xc})^2\Delta r_{yc} \\ & + 9\rho_{xc}\rho_{yc}(\Delta r_{xc})^2(\Delta r_{yc})^2;\end{aligned}$$

$$\begin{aligned}(r_{xc}r_{yc})^4 = {} & \rho^4_{xc}\rho^4_{yc} + \rho^4_{xc}(\Delta r_{yc})^4 + \rho^4_{yc}(\Delta r_{xc})^4 + (\Delta r_{xc})^4(\Delta r_{yc})^4 + 4\rho^4_{xc}\rho^3_{yc}\Delta r_{yc} \\ & + 4\rho^3_{xc}\rho^4_{yc}\Delta r_{xc} + 4\rho^3_{yc}(\Delta r_{xc})^4\Delta r_{yc} + 4\rho_{xc}(\Delta r_{xc})^3(\Delta r_{yc})^4 + 4\rho^4_{xc}\rho_{yc}(\Delta r_{yc})^3 \\ & + 4\rho^3_{xc}\Delta r_{xc}(\Delta r_{yc})^4 + 4\rho_{xc}\rho^4_{yc}(\Delta r_{xc})^3 + 4\rho_{yc}(\Delta r_{xc})^4(\Delta r_{yc})^3 + 6\rho^4_{xc}\rho^2_{yc}(\Delta r_{yc})^2 \\ & + 6\rho^2_{xc}\rho^4_{yc}(\Delta r_{xc})^2 + 6\rho^2_{xc}(\Delta r_{xc})^2(\Delta r_{yc})^4 + 6\rho^2_{yc}(\Delta r_{xc})^4(\Delta r_{yc})^2 \\ & + 16\rho^3_{xc}\rho^3_{yc}\Delta r_{xc}\Delta r_{yc} + 16\rho^3_{xc}\rho_{yc}\Delta r_{xc}(\Delta r_{yc})^3 + 16\rho_{xc}\rho^3_{yc}(\Delta r_{xc})^3\Delta r_{yc} \\ & + 16\rho_{xc}\rho_{yc}(\Delta r_{xc})^3(\Delta r_{yc})^3 + 24\rho^3_{xc}\rho^2_{yc}\Delta r_{xc}(\Delta r_{yc})^2 + 24\rho^2_{xc}\rho^3_{yc}(\Delta r_{xc})^2\Delta r_{yc} \\ & + 24\rho^2_{xc}\rho_{yc}(\Delta r_{xc})^2(\Delta r_{yc})^3 + 24\rho_{xc}\rho^2_{yc}(\Delta r_{xc})^3(\Delta r_{yc})^2 \\ & + 36\rho^2_{xc}\rho^2_{yc}(\Delta r_{xc})^2(\Delta r_{yc})^2.\end{aligned}$$

Dropping terms of order higher than the fourth regarding the deviations Δr_{xc} and Δr_{yc} in the above expressions, and taking expectations of them yield

$$
\begin{aligned}
&E(r_{xc}r_{yc}) = \rho_{xc}\rho_{yc} + E(\Delta r_{xc}\Delta r_{yc});\\
&E(r_{xc}r_{yc})^2 = \rho^2_{xc}\rho^2_{yc} + \rho^2_{xc}E(\Delta r_{yc})^2 + \rho^2_{yc}E(\Delta r_{xc})^2 + E[(\Delta r_{xc})^2(\Delta r_{yc})^2]\\
&\qquad + 2\rho_{yc}E[(\Delta r_{xc})^2\Delta r_{yc}] + 2\rho_{xc}E[\Delta r_{xc}(\Delta r_{yc})^2] + 4\rho_{xc}\rho_{yc}E(\Delta r_{xc}\Delta r_{yc});\\
&E(r_{xc}r_{yc})^3 = \rho^3_{xc}\rho^3_{yc} + \rho^3_{xc}E(\Delta r_{yc})^3 + \rho^3_{yc}E(\Delta r_{xc})^3 + 3\rho^3_{xc}\rho_{yc}E(\Delta r_{yc})^2\\
&\qquad + 3\rho_{xc}\rho^3_{yc}E(\Delta r_{xc})^2 + 3\rho^2_{xc}E[\Delta r_{xc}(\Delta r_{yc})^3] + 3\rho^2_{yc}E[(\Delta r_{xc})^3\Delta r_{yc}]\\
&\qquad + 9\rho^2_{xc}\rho^2_{yc}E(\Delta r_{xc}\Delta r_{yc}) + 9\rho^2_{xc}\rho_{yc}E[\Delta r_{xc}(\Delta r_{yc})^2]\\
&\qquad + 9\rho_{xc}\rho^2_{yc}E[(\Delta r_{xc})^2\Delta r_{yc}] + 9\rho_{xc}\rho_{yc}E[(\Delta r_{xc})^2(\Delta r_{yc})^2];\\
&E(r_{xc}r_{yc})^4 = \rho^4_{xc}\rho^4_{yc} + \rho^4_{xc}E(\Delta r_{yc})^4 + \rho^4_{yc}E(\Delta r_{xc})^4 + 4\rho^4_{xc}\rho_{yc}E(\Delta r_{yc})^3\\
&\qquad + 4\rho_{xc}\rho^4_{yc}E(\Delta r_{xc})^3 + 6\rho^4_{xc}\rho^2_{yc}E(\Delta r_{yc})^2 + 6\rho^2_{xc}\rho^4_{yc}E(\Delta r_{xc})^2\\
&\qquad + 16\rho^3_{xc}\rho^3_{yc}E(\Delta r_{xc}\Delta r_{yc}) + 16\rho^3_{xc}\rho_{yc}E[\Delta r_{xc}(\Delta r_{yc})^3]\\
&\qquad + 16\rho_{xc}\rho^3_{yc}E[(\Delta r_{xc})^3\Delta r_{yc}] + 24\rho^3_{xc}\rho^2_{yc}E[\Delta r_{xc}(\Delta r_{yc})^2]\\
&\qquad + 24\rho^2_{xc}\rho^3_{yc}E[(\Delta r_{xc})^2\Delta r_{yc}] + 36\rho^2_{xc}\rho^2_{yc}E[(\Delta r_{xc})^2(\Delta r_{yc})^2].
\end{aligned}
\qquad (4)
$$

Note that $E(\Delta r_{xc}) = E(r_{xc} - \rho_{xc}) = E(r_{xc}) - \rho_{xc} = 0$, so as $E(\Delta r_{yc}) = 0$.

Suppose the three initial variables X, Y, and C follow a trivariate normal distribution. Then, we get the asymptotic variances and covariance of r_{xc} and r_{yc} as follows (Hotelling, 1936, 1940):

$$\left.\begin{aligned} E(\Delta r_{xc})^2 &= \frac{1}{N-1}(1-\rho_{xc}^2)^2 := \sigma_{xc}^{(2)}; \\ E(\Delta r_{yc})^2 &= \frac{1}{N-1}(1-\rho_{yc}^2)^2 := \sigma_{yc}^{(2)}; \\ E(\Delta r_{xc}\Delta r_{yc}) &= \frac{1}{N-1}\left[\rho_{xy}(1-\rho_{xc}^2-\rho_{yc}^2)-\frac{1}{2}\rho_{xc}\rho_{yc}(1-\rho_{xc}^2-\rho_{yc}^2-\rho_{xy}^2)\right] := \sigma_{xc\,yc}. \end{aligned}\right\} \quad (5)$$

where N is the sample size and ":=" means "defined as".

From Ghosh (1966), we know that $E(\Delta r_{xc})^3$, $E(\Delta r_{yc})^3$, $E(\Delta r_{xc})^4$, and $E(\Delta r_{yc})^4$ can be approximately expressed as follows

$$\left.\begin{aligned} E(\Delta r_{xc})^3 &= -\frac{6}{(N-1)^2}\rho_{xc}(1-\rho_{xc}^2)^3 := \sigma_{xc}^{(3)}; \\ E(\Delta r_{yc})^3 &= -\frac{6}{(N-1)^2}\rho_{yc}(1-\rho_{yc}^2)^3 := \sigma_{yc}^{(3)}; \\ E(\Delta r_{xc})^4 &= \frac{3}{(N-1)^2}(1-\rho_{xc}^2)^4 = 3(\sigma_{xc}^{(2)})^2; \\ E(\Delta r_{yc})^4 &= \frac{3}{(N-1)^2}(1-\rho_{yc}^2)^4 = 3(\sigma_{yc}^{(2)})^2. \end{aligned}\right\} \quad (6)$$

Note that originally in Ghosh (1966) the denominator was $m = N + 6$. To be consistent with Hotelling's formulae we continue to use $N - 1$ instead of $N + 6$ in (6).

The calculation of the approximations to the expectations of the cross products of Δr_{xc} and Δr_{yc} can be carried out in the following manner. According to the asymptotic distribution theory, the distribution of a set of correlation coefficients approaches multivariate normal, as the sample size becomes large enough (Steiger, 1980, p.246). Under this condition, the expected value of the third order of the cross products become asymptotically vanished, that is, $E[(\Delta r_{xc})^2 \Delta r_{yc}] = E[\Delta r_{xc}(\Delta r_{yc})^2] = 0$ asymptotically, and

$$E(\Delta r_i \Delta r_j \Delta r_k \Delta r_l) = E(\Delta r_i \Delta r_j)E(\Delta r_k \Delta r_l) + E(\Delta r_i \Delta r_k)E(\Delta r_j \Delta r_l) + E(\Delta r_i \Delta r_l)E(\Delta r_j \Delta r_k). \quad (7)$$

where i, j, k, and l can be xc or yc (see Anderson, 1958, p. 39).

If we set $i = j = k = xc$, $l = yc$, and $i = j = k = yc$, $l = xc$, respectively, then (7) yields

$$E[(\Delta r_{xc})^3(\Delta r_{yc})] = 3E(\Delta r_{xc})^2 E(\Delta r_{xc}\Delta r_{yc}) = 3\sigma^{(2)}_{xc}\sigma_{xc\,yc}, \quad (8)$$

$$E[(\Delta r_{yc})^3(\Delta r_{xc})] = 3E(\Delta r_{yc})^2 E(\Delta r_{xc}\Delta r_{yc}) = 3\sigma^{(2)}_{yc}\sigma_{xc\,yc}. \quad (9)$$

If we set $i = j = xc$ and $k = l = yc$, then (7) yields

$$E[(\Delta r_{yc})^2(\Delta r_{xc})^2] = E(\Delta r_{xc})^2 E(\Delta r_{yc})^2 + 2[E(\Delta r_{xc}\Delta r_{yc})]^2 = \sigma^{(2)}_{xc}\sigma^{(2)}_{yc} + 2\sigma^2_{xc\,yc}. \quad (10)$$

Applying (5), (6), (8), (9), and (10) to (4), we get the approximate first four non-central moment of $r_{xc}r_{yc}$ as follows:

$$\mu'_1 = E(r_{xc}r_{yc}) = \rho_{xc}\rho_{yc} + \sigma_{xc\,yc};$$

$$\mu'_2 = E(r_{xc}r_{yc})^2 = \rho^2_{xc}\rho^2_{yc} + \rho^2_{xc}\sigma^{(2)}_{yc} + \rho^2_{yc}\sigma^{(2)}_{xc} + \sigma^{(2)}_{xc}\sigma^{(2)}_{yc} + 2\sigma^2_{xc\,yc}$$

$$+ 4\rho_{xc}\rho_{yc}\sigma_{xc\,yc};$$

$$\mu'_3 = E(r_{xc}r_{yc})^3 = \rho^3_{xc}\rho^3_{yc} + \rho^3_{xc}\sigma^{(3)}_{yc} + \rho^3_{yc}\sigma^{(3)}_{xc} + 3\rho^3_{xc}\rho_{yc}\sigma^{(2)}_{yc} + 3\rho_{xc}\rho^3_{yc}\sigma^{(2)}_{xc}$$
$$+ 9\rho^2_{xc}\sigma^{(2)}_{yc}\sigma_{xc\,yc} + 9\rho^2_{yc}\sigma^{(2)}_{xc}\sigma_{xc\,yc} + 9\rho^2_{xc}\rho^2_{yc}\sigma_{xc\,yc}$$
$$+ 9\rho_{xc}\rho_{yc}\sigma^{(2)}_{xc}\sigma^{(2)}_{yc} + 18\rho_{xc}\rho_{yc}\sigma^2_{xc\,yc};$$
$$\mu'_4 = E(r_{xc}r_{yc})^4 = \rho^4_{xc}\rho^4_{yc} + 3\rho^4_{xc}(\sigma^{(2)}_{yc})^2 + 3\rho^4_{yc}(\sigma^{(2)}_{xc})^2 + 4\rho^4_{xc}\rho_{yc}\sigma^{(3)}_{yc}$$
$$+ 4\rho_{xc}\rho^4_{yc}\sigma^{(3)}_{xc} + 6\rho^4_{xc}\rho^2_{yc}\sigma^{(2)}_{yc} + 6\rho^2_{xc}\rho^4_{yc}\sigma^{(2)}_{xc} + 16\rho^3_{xc}\rho^3_{yc}\sigma_{xc\,yc}$$
$$+ 48\rho^3_{xc}\rho_{yc}\sigma^{(2)}_{yc}\sigma_{xc\,yc} + 48\rho_{xc}\rho^3_{yc}\sigma^{(2)}_{xc}\sigma_{xc\,yc} + 36\rho^2_{xc}\rho^2_{yc}\sigma^{(2)}_{xc}\sigma^{(2)}_{yc}$$
$$+ 72\rho^2_{xc}\rho^2_{yc}\sigma^2_{xc\,yc}.$$

For the central moments about the mean, substitution in Kendall and Stuart`s equations (Kendall & Stuart, 1977, equation (3.9), p. 58) gives

$$\left.\begin{aligned}
\mu_2 &= \mu'_2 - \mu'^2_1 \\
&= \rho^2_{yc}\sigma^{(2)}_{xc} + \rho^2_{xc}\sigma^{(2)}_{yc} + \sigma^{(2)}_{xc}\sigma^{(2)}_{yc} + 2\rho_{xc}\rho_{yc}\sigma_{xc\,yc} + \sigma^2_{xc\,yc}; \\
\mu_3 &= \mu'_3 - 3\mu'_1\mu'_2 + 2\mu'^3_1 \\
&= 6\rho_{xc}\rho_{yc}\sigma^{(2)}_{xc}\sigma^{(2)}_{yc} + \rho^3_{yc}\sigma^{(3)}_{xc} + \rho^3_{xc}\sigma^{(3)}_{yc} + 6\rho^2_{yc}\sigma^{(2)}_{xc}\sigma_{xc\,yc} \\
&\quad + 6\rho^2_{xc}\sigma^{(2)}_{yc}\sigma_{xc\,yc} - 3\sigma^{(2)}_{xc}\sigma^{(2)}_{yc}\sigma_{xc\,yc} + 6\rho_{xc}\rho_{yc}\sigma^2_{xc\,yc} - 4\sigma^3_{xc\,yc}; \\
\mu_4 &= \mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1 \\
&= 3\rho^4_{yc}\sigma^{(2)}_{xc} + 6\rho^2_{xc}\rho^2_{yc}\sigma^{(2)}_{xc}\sigma^{(2)}_{yc} + 3\rho^4_{xc}\sigma^{(2)}_{yc} + 12\rho_{xc}\rho^3_{yc}\sigma^{(2)}_{xc}\sigma_{xc\,yc} \\
&\quad + 12\rho^3_{xc}\rho_{yc}\sigma^{(2)}_{yc}\sigma_{xc\,yc} - 24\rho_{xc}\rho_{yc}\sigma^{(2)}_{xc}\sigma^{(2)}_{yc}\sigma_{xc\,yc} - 4\rho^3_{yc}\sigma^{(3)}_{xc}\sigma_{xc\,yc} \\
&\quad - 4\rho^3_{xc}\sigma^{(3)}_{yc}\sigma_{xc\,yc} + 12\sigma^{(2)}_{xc}\sigma^{(2)}_{yc}\sigma^2_{xc\,yc} - 30\rho^2_{yc}\sigma^{(2)}_{xc}\sigma^2_{xc\,yc} \\
&\quad - 30\rho^2_{xc}\sigma^{(2)}_{yc}\sigma^2_{xc\,yc} + 6\sigma^{(2)}_{xc}\sigma^{(2)}_{yc}\sigma^2_{xc\,yc} - 36\rho_{xc}\rho_{yc}\sigma^3_{xc\,yc} + 9\sigma^4_{xc\,yc};
\end{aligned}\right\} \quad (7)$$

Note that $\mu_1 = \int_{-\infty}^{\infty}(r_{xc}r_{yc} - \mu'_1)dF = \int_{-\infty}^{\infty}[r_{xc}r_{yc} - E(r_{xc}r_{yc})]dF = \int_{-\infty}^{\infty}(r_{xc}r_{yc})dF - E(r_{xc}r_{yc}) = 0$.

Discussion

As the moments of the product of the two dependent correlations are approximate, a simulation study was conducted to test the accuracy of the approximation method in this study.

The parameters in this simulation study are the sample size, N, and the population correlations, ρ_{xc}, ρ_{yc}, and ρ_{xy}. For ρ_{xc}, ρ_{yc}, and ρ_{xy}, we chose .10, .30, and .50 as small, medium, and large correlations (Cohen, 1988); and for the sample size, we selected 28, 84, and 783, which give us a statistical power of .80 for the small, medium, and large sizes of correlations (Cohen & Cohen, 1983). We did not include every possible combination of .10, .30, and .50. It is because ρ_{xc} and ρ_{yc} are symmetric in the mathematical expressions in the approximation section. We also removed the mirror ones. In addition, under the condition of $\rho_{xc}\rho_{yc}\rho_{xy} \geq 0$, we could always change the sign of the relevant variable to have the all three correlations positive. Thus, there are only 18 positive correlation triplets left for this simulation study. By the three magnitudes of N, we totally have $18 \times 3 = 54$ cells to simulate.

For each of 54 cells (N, ρ_{xc}, ρ_{yc}, ρ_{xy}), we, by Cholesky factorization, generated X, Y, and C with the specified population correlations (see, Browne, 1968) and with the sample size as N. For simplicity, we only generated (X, Y, C) as a trivariate standard normal distribution. Then, we computed a pair of estimated correlation coefficients r_{xc}

and r_{yc}. Within each cell, the same procedure was replicated 1000 times, getting 1000 pairs of r_{xc} and r_{yc}. Directly multiplying r_{xc} and r_{yc} gave us a simulated distribution of the product of $r_{xc}r_{yc}$, which serves as the true distribution of $r_{xc}r_{yc}$, with which the approximate moments are compared. Table 1 shows the comparison results for the moments (mean, variance, skewness, and kurtosis).

Insert Table 1 about here

From Table 1 we can see that the approximated mean, variance, and skewness are accurate enough to the simulated ones, but the kurtosis is not. It is probably due to the approximation to the expected values of the cross products of order higher than the second regarding the deviations Δr_{xc} and Δr_{yc} during the derivations. In this approximation method, the asymptotic normal distribution theory is applied to correlation coefficients, so a very large sample size is needed. Here, our sample size is only 28, 84, or 783. If the sample size were increased from 783 to10,000, the error of the approximated kurtosis would be reduced from 1.5654 to .2623 for $\rho_{xc} = .1$, $\rho_{yc} = .1$, and $\rho_{xy} = .5$, which is still not good enough. Thus, in general the simulation results are not very pleasing. To get more accurate approximated kurtosis will be one of the main tasks in the further research.

As stated in the introduction section, the distribution of the product of two correlations may be applicable to other social science studies. For example, the distribution of the product could be applied to assessing indirect effects with respect to mediating vari-

ables in causal inference. This is because the indirect effect is kind of product of two correlations. Thus, if we know the distribution of the product, we could understand the indirect effect better. Also, the distribution of the product may have potential applications to generalizability theory.

We know that it is not necessary for predictors and covariates to be normal distributed in linear models. Hence, for further study, it would be interesting to find a distribution of the product of two dependent correlations for categorical or mixed initial variables, X, Y, and C.

TABLE 1.
Simulation results

			MEAN		VARIANCE		SKEWNESS		KURTOSIS	
ρ_{xc}	ρ_{yc}	ρ_{xy}	Approx.	Simul.	Approx.	Simul.	Approx.	Simul.	Approx.	Simul.
					$N = 28$					
0.1	0.1	0.1	0.01345	**0.01624**	0.00213	**0.00201**	0.81083	**1.14098**	-2.85635	**3.30262**
		0.3	0.02072	**0.02271**	0.00237	**0.00220**	0.72444	**1.54119**	-3.45987	**3.94398**
		0.5	0.02801	**0.02906**	0.00273	**0.00244**	0.56703	**1.77346**	-4.03432	**4.36679**
	0.3	0.1	0.03284	**0.03579**	0.00487	**0.00471**	0.67691	**0.71170**	-1.34367	**1.13692**
		0.3	0.03955	**0.04186**	0.00535	**0.00508**	0.92676	**0.94735**	-1.78507	**1.24426**
		0.5	0.04631	**0.04786**	0.00593	**0.00550**	1.10041	**1.13062**	-2.43586	**1.39534**
	0.5	0.1	0.05206	**0.05482**	0.01025	**0.01003**	0.22826	**0.31122**	-0.44977	**0.30261**
		0.3	0.05762	**0.05994**	0.01086	**0.01051**	0.48435	**0.49120**	-0.60683	**0.31766**
		0.5	0.06325	**0.06504**	0.01154	**0.01102**	0.70714	**0.65820**	-0.90530	**0.43916**
0.3	0.3	0.1	0.09169	**0.09383**	0.00677	**0.00682**	0.82642	**0.70989**	-0.84848	**0.27878**
		0.3	0.09789	**0.09950**	0.00795	**0.00759**	0.95800	**0.77837**	-1.09620	**0.26495**
		0.5	0.10424	**0.10508**	0.00923	**0.00836**	1.05699	**0.83423**	-1.45629	**0.28237**
	0.5	0.1	0.15064	**0.15152**	0.01037	**0.01045**	0.31359	**0.31089**	-0.36652	**-0.01671**
		0.3	0.15575	**0.15634**	0.01194	**0.01151**	0.49271	**0.39215**	-0.44854	**-0.14494**
		0.5	0.16108	**0.16116**	0.01363	**0.01260**	0.63731	**0.46428**	-0.62217	**-0.20303**
0.5	0.5	0.1	0.24958	**0.24874**	0.01064	**0.01120**	0.17467	**0.15733**	-0.23707	**-0.21488**
		0.3	0.25366	**0.25292**	0.01269	**0.01272**	0.32218	**0.15577**	-0.25593	**-0.32563**
		0.5	0.25810	**0.25713**	0.01497	**0.01419**	0.44190	**0.16676**	-0.34035	**-0.42805**
					$N = 84$					
0.1	0.1	0.1	0.01112	**0.01040**	0.00040	**0.00038**	1.17654	**1.40298**	-2.02490	**4.22240**
		0.3	0.01349	**0.01265**	0.00046	**0.00043**	1.25079	**1.75350**	-2.63499	**5.21923**
		0.5	0.01586	**0.01491**	0.00053	**0.00048**	1.26456	**1.98555**	-3.33793	**5.96288**
	0.3	0.1	0.03092	**0.03021**	0.00134	**0.00136**	0.51147	**0.64165**	-0.56455	**1.30761**
		0.3	0.03311	**0.03232**	0.00148	**0.00147**	0.72127	**0.84421**	-0.76818	**1.62047**
		0.5	0.03530	**0.03450**	0.00163	**0.00159**	0.88886	**1.01121**	-1.08581	**1.93131**
	0.5	0.1	0.05067	**0.05055**	0.00317	**0.00332**	0.14250	**0.26849**	-0.15817	**0.37118**
		0.3	0.05248	**0.05236**	0.00335	**0.00348**	0.30563	**0.40076**	-0.21550	**0.48633**
		0.5	0.05431	**0.05434**	0.00355	**0.00366**	0.45196	**0.52682**	-0.32659	**0.61502**
0.3	0.3	0.1	0.09055	**0.08955**	0.00200	**0.00206**	0.55046	**0.56488**	-0.32080	**0.59579**
		0.3	0.09257	**0.09130**	0.00236	**0.00236**	0.63952	**0.63146**	-0.41516	**0.66301**
		0.5	0.09463	**0.09307**	0.00275	**0.00266**	0.71281	**0.67737**	-0.55673	**0.67062**
	0.5	0.1	0.15021	**0.14950**	0.00323	**0.00349**	0.19110	**0.26190**	-0.12708	**0.19927**
		0.3	0.15187	**0.15092**	0.00374	**0.00392**	0.30151	**0.30654**	-0.15563	**0.20749**
		0.5	0.15361	**0.15245**	0.00427	**0.00437**	0.39190	**0.34178**	-0.21712	**0.19508**
0.5	0.5	0.1	0.24986	**0.24891**	0.00337	**0.00377**	0.10369	**0.11715**	-0.08039	**0.06520**
		0.3	0.25119	**0.24998**	0.00403	**0.00436**	0.19189	**0.10792**	-0.08655	**0.03483**
		0.5	0.25264	**0.25104**	0.00476	**0.00496**	0.26356	**0.11800**	-0.11533	**-0.02066**
					$N = 783$					
0.1	0.1	0.1	0.01012	**0.01004**	0.00003	**0.00003**	0.70661	**0.72713**	-0.38403	**0.71404**
		0.3	0.01037	**0.01029**	0.00003	**0.00003**	0.77223	**0.79460**	-0.51204	**0.82029**
		0.5	0.01062	**0.01055**	0.00004	**0.00004**	0.82774	**0.84066**	-0.67696	**0.88843**
	0.3	0.1	0.03010	**0.03002**	0.00013	**0.00013**	0.19086	**0.08697**	-0.06809	**0.21015**
		0.3	0.03033	**0.03025**	0.00015	**0.00014**	0.27212	**0.17025**	-0.09364	**0.18834**
		0.5	0.03056	**0.03050**	0.00016	**0.00016**	0.34070	**0.23844**	-0.13466	**0.18065**
	0.5	0.1	0.05007	**0.04997**	0.00033	**0.00032**	0.04833	**-0.03504**	-0.01739	**0.14316**
		0.3	0.05026	**0.05017**	0.00035	**0.00034**	0.10412	**0.02031**	-0.02379	**0.12511**
		0.5	0.05046	**0.05037**	0.00037	**0.00036**	0.15484	**0.07422**	-0.03632	**0.10797**
0.3	0.3	0.1	0.09006	**0.08968**	0.00020	**0.00020**	0.19266	**0.22054**	-0.03654	**0.01582**
		0.3	0.09027	**0.08987**	0.00024	**0.00023**	0.22380	**0.22524**	-0.04728	**0.06319**
		0.5	0.09049	**0.09009**	0.00028	**0.00027**	0.25044	**0.22215**	-0.06366	**0.08653**
	0.5	0.1	0.15002	**0.14962**	0.00034	**0.00033**	0.06412	**0.02306**	-0.01388	**0.08448**
		0.3	0.15020	**0.14979**	0.00039	**0.00038**	0.10132	**0.03717**	-0.01700	**0.09751**
		0.5	0.15038	**0.14999**	0.00045	**0.00043**	0.13197	**0.04316**	-0.02378	**0.10608**
0.5	0.5	0.1	0.24999	**0.24935**	0.00035	**0.00034**	0.03438	**0.08701**	-0.00869	**-0.03908**
		0.3	0.25013	**0.24941**	0.00042	**0.00041**	0.06371	**0.08605**	-0.00935	**-0.00648**
		0.5	0.25028	**0.24952**	0.00050	**0.00048**	0.08755	**0.07288**	-0.01246	**0.01104**

16 BEST COPY AVAILABLE

References

Abbott, A. (1998). The casual devolution. *Sociological Methods and Research, 27*(2), 148-181.

Anderson, T. W. (1958). *Introduction to multivariate statistical analysis.* New York: John Wiley and Sons.

Aroian, L. A. (1947). The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics, 18*(2), 265-271.

Aroian, L. A., Taneja, V. S., & Cornwell, L. W. (1978). Mathematical forms of the distribution of the product of two normal variables. *Communications in Statistics: Part A—Theory and Methods, 7*(2), 165-172.

Boyer, J. E., Jr., Palachek, A. D., & Schucany, W. R. (1983). An empirical study of related correlation coefficients. *Journal of Educational Statistics, 8*(1), 75-86.

Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, 33, 267-334.

Choi, S. C. (1977). Tests of equality of dependent correlation coefficients. *Biometrika, 64*(3), 645-647.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., & Cohen, P. (1983). *Applied multivariate regression/correlation analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Cook, T., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Cornwell, L. W., Aroian, L. A., & Taneja, V. S. (1978). Numerical evaluation of the distribution of the product of two normal variables. *Journal of Statistical Computing and Simulation, 7*, 123-131.

Craig, C. C. (1936). On the frequency function of *xy*. *The Annals of Mathematical Statistics, 7*(1), 1-15.

Craig, C. C. (1942). On frequency distributions of the quotient and of the product of two statistical variables. *American Mathematical Monthly, 49*(1), 24-32.

Dunn, O. J., & Clark, V. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association, 64*, 366-377.

Dunn, O. J., & Clark, V. (1971). Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association, 66*, 904-908.

Frank, K. A. (In press). The impact of a confounding variable on a regression coefficient. To appear in *Sociological Methods and Research.*

Ghosh, B. K. (1966). Asymptotic expansions for the moments of the distribution of correlation coefficient. *Biometricka, 35*, 258-262.

Hotelling, H. (1936). Relations between two sets of variables. *Biometricka, 28*, 321-377.

Hotelling, H. (1940). The selection of variables for use in prediction with some comments on the general problem of nuisance parameters. *The Annals of Mathematical Statistics, 11*(3), 271-283.

Kendall, M., Sir, & Stuart, A. (1977). *The advanced theory of statistics, Vol.1, Distribution theory* (4th ed.). New York: Macmillan.

Mathai, A. M., & Saxena, R. K. (1969). Distribution of a product and the structural set up of densities. *The Annals of Mathematical Statistics, 40*(4), 1439-1448.

Mathai, A. M., & Saxena, R. K. (1978). *The* H-*function with applications in statistics and other disciplines*. New York: John Wiley & Sons.

May, K., & Hittner, J. B. (1997a). A note on statistics for comparing dependent correlations. *Psychological Reports, 80*, 475-480.

May, K., & Hittner, J. B. (1997b). Tests for comparing dependent correlations revisited: A Monte Carlo study. *Journal of Experimental Education, 65*(3), 257-269.

Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). The product of two normally distributed random variables. In W. Kenney & R. Odeh (Eds.), *Selected tables in mathematical statistics* (Vol. VII, pp. 1-256). Providence, IR: American Mathematical Society.

Meeker, W. Q., & Escobar, L. A. (1994). An algorithm to compute the CDF of the product of two normal random variables. *Communications in Statistics: Part B—Simulation and Computation, 23*(1), 271-280.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*(1), 172-175.

NCES (1992). *User's manual: NELS:88 first follow-up student component volume 1*. Washington, DC: U.S. Government Printing Office.

Neill, J. J., & Dunn, O. J. (1975). Equality of dependent correlation coefficients. *Biometrics, 31*, 531-543.

Sobel, M. E. (1996). An introduction to casual inference. *Sociological Methods and Research, 24*(3), 353-379.

Springer, M. D. (1983). [Review of *Selected tables in mathematical statistics (Vol. VII), The product of two normally distributed random variables*]. *Technometrics, 25*(2), 211.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245-251.

Wallgren, C. M. (1980). The distribution of the product of two correlated *t* variables. *Journal of the American Statistical Association, 75*, 996-1000.

Williams, E. J. (1959). Significance of difference between two non-independent correlation coefficients. *Biometrics, 15*, 135-136.

Wolfe, D. A. (1976). On testing equality of related correlation coefficients. *Biometrika, 63*, 214-215.